The Influence of Speech Enhancement Algorithm in Speech Compression with Voice Excited Linear Predictive Coding

D. Deepa*[1], C. Poongodi[2], Dr. A. Shanmugam[3]

Bannari Amman Institute of Technology, Sathyamangalam, Tamilnadu, India

*[1]deepa_dhanaskodi@yahoo.co.in; [2] poongi_cj@yahoo.co.in; [3]dras@yahoo.com

Abstract

Problem statement: Speech Enhancement plays an important role in any of the speech processing systems like speech recognition, speech coding, mobile communication, hearing aid, etc., **Approach:** In this work, the performance of the speech coding method is enhanced by using speech enhancement as the preprocessing technique. The purpose of the proposed method is to reduce the bit rate of the speech signal to be transmitted, so that the bandwidth can be utilized efficiently. In noisy environment speech coding is done both for desired speech and the unwanted noise signal. If the noise is reduced before coding the speech signal, the bit rate required will also be reduced. In this work a simple adaptive speech enhancement technique, using an adaptive sigmoid type function to determine the weighting factor of the TSDD algorithm is employed based on a subband approach for speech enhancement and Voice excited Linear predictive coding (VELP) method is used for coding the speech signal. **Results:** Objective and subjective measures like SNR and MSE were obtained, which shows the ability of the proposed method for efficient coding of the speech signal with noise removal. **Conclusion/Recommendations:** Performance assessment shows that our proposal can achieve a more significant noise reduction and a better coding of the speech signal as compared to the conventional methods.

Keywords

Subband Two Step Decision Directed Approach; Posteriori SNR; Mean Square Error (MSE); Adaptive Weighting Factor; Global Masking Threshold; Signal-To Noise Ratio (SNR); Voice Excited Linear Predictive Coding Technique; Bit Rate

Introduction

Speech coding has been and still is a major issue in the area of digital speech processing. There exist many types of speech compression that take use of different techniques. However, most methods of speech compression exploit the fact that speech production occurs through slow anatomical movements and that the speech produced has a limited frequency range. The frequency of human speech ranges from around 300 Hz to 3400 Hz. Most forms of speech coding are usually based on a lossy algorithm. Lossy algorithms are considered acceptable when encoding the speech signal, because some amount of loss in quality is often undetectable to the human ear.

Another fact about speech production that can be taken advantage of is that mechanically there is a high correlation between adjacent samples of speech. Most forms of speech compression are achieved by modeling the process of speech production as a linear digital filter. There present many other characteristics about speech production that can be exploited by speech coding algorithms. One fact that is often used is that the period of silence takes up greater than 50% of conversations. An easy way to save bandwidth and reduce the amount of information needed to represent the speech signal is not to transmit the silence.

All vocoders, including LPC (Linear Predictive Coding) vocoders, have four main attributes: bit rate, delay, complexity, quality. Any voice coder, regardless of the algorithm it uses, will have to make tradeoffs between these different attributes. First attribute of vocoders of the bit rate, is used to determine the degree of compression that a vocoder achieves. Uncompressed speech is usually transmitted at 64 kb/s using 8 bits/sample and a rate of 8 kHz for sampling. Any bit rate below 64 kb/s is considered compression. The linear predictive coder transmits speech at a bit rate of 2.4 kb/s, an excellent rate of compression. This parameter is mainly considered in this work. Delay is another important attribute for vocoders that are involved with the transmission of an encoded speech signal. Vocoders which are involved with the storage of the compressed speech, as opposed to transmission, are not as concern with delay. The general delay standard for transmitted speech conversations is that any delay that is greater than 300 ms is considered unacceptable. The third attribute of voice coders is the complexity of the algorithm used. The complexity affects both the cost and the power of the vocoder. Linear predictive coding because of its high compression rate is very complex and involves executing millions of instructions per second. LPC often requires more than one processor to run in real time. The final attribute of vocoders is quality. Quality is a subjective attribute and it depends on how the speech sounds to a given listener. This part of attribute is also considered in this work and the preprocessing of the speech signal with speech enhancement will further improve the quality, which is tested by different subjective and objective quality measures. This test involves subjects being given pairs of sentences and asked to rate them as excellent, good, fair, poor, or bad. The speech coder that is developed is analyzed using both subjective and objective analysis. Subjective analysis will consist of listening to the encoded speech Signal and making judgments on its quality. The quality of played back speech will be solely based on the opinion of the listener. An objective analysis will be performed by computing Signal to Noise Ratio (SNR), Mean square error (MSE) and IS distance measure between the original and the coded speech signal.

LPC Model

LPC model has two key components, analysis and synthesis or encoding and decoding part. In the analysis part the speech signal is segmented into blocks. The principle behind the use of LPC is to minimize the difference between the original and estimated signal over a finite duration. It will give a unique set of coefficients which is estimated in every frame. This segments will be normally of 20 ms.

An easy way to comply with the journal paper formatting requirements is to use this document as a template and simply type your text into it.

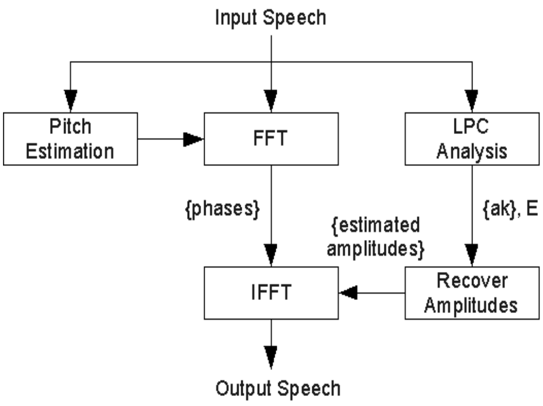

FIGURE 1 BLOCK DIAGRAM OF LPC MODEL

The main idea behind the voice-excitation is to avoid the imprecise detection of the pitch and the use of an impulse train while synthesizing the speech. Thus the Input speech signal in each frame is filtered with the estimated transfer function of LPC analyzer. The filtered signal is called residual. If this signal is transmitted to the receiver one can achieve a very good quality. For a good reconstruction of the excitation only the low frequencies of the residual signals are coded.

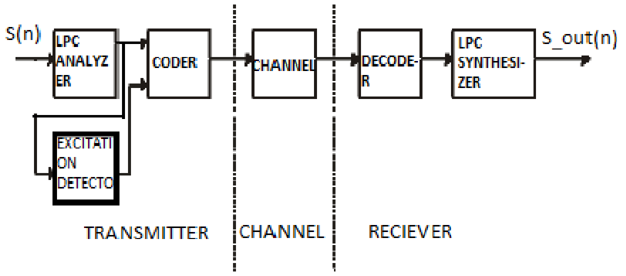

FIGURE 2 BLOCK DIAGRAM OF A VOICE EXCITED LPC VOCODER

Figure 1 and 2 give the block diagram for LPC model and VELP model. To achieve a high compression rate we have employed the Discrete cosine transform (DCT) of the residual signal. Thus one way to compress the signal is to transfer only the coefficients, which contains most of the energy.

Speech Enhancement Technique

Most of the methods for enhancing the speech signal have been developed with some or other auditory, perceptual or statistical constraints placed on the speech and noise signals. However, in real situations, it is very difficult to reliably predict the characteristics of the interfering noise signal or the exact characteristics of the speech waveform. Hence, in effect, the speech enhancement methods are sub optimal and only reduce the amount of noise in the signal to some extent. Due to this nature, some of the speech signals are distorted during the process. The best tradeoff between speech distortion and noise reduction in a perceptual sense are based on properties closely related to human perception and masking effects. Masking is a fundamental aspect of the human auditory system and is a basic element of perceptual coding systems which is utilized in enhancement systems.

In the proposed method before speech coding input noisy speech signal is enhanced using Subband two step decision directed approach with adaptive noise estimation, it is the frequency dependent processing of the two step decision directed procedure. This method offers a better quality of the enhanced speech with reduced residual noise. The block diagram of the method is given in Figure 6. This approach has been justified due to variation in signal to noise ratio across the speech spectrum. The noise spectrum does not affect the speech signal uniformly over the whole spectrum some frequencies are affected more adversely than others. To take into account the fact that colored noise affects the speech spectrum differently at various frequencies, this subband approach is applied to the two step decision directed method. The speech spectrum is divided into eight non overlapping bands, and TSDD approach is performed independently in each band. The estimate of the clean speech in the ith band is obtained by:

$$\left|\hat{\mathrm{S}}_{\mathbf{i}}(\mathrm{w})\right| = g_{i}^{TSDD}.\left|\mathrm{Y}(\mathrm{w})\right|_{;}\ \mathbf{b}_{\mathbf{i}} \le \mathbf{w} \le \mathbf{e}_{\mathbf{i}}$$

where b_i and e_i are the beginning and ending frequency bins of the ith frequency.

Although the amount of residual noise is greatly reduced with the TSDD approach, some small amount of musical noise and speech distortion is found to be present in the estimated clean speech. In order to further enhance the quality of speech, a psychoacoustically motivated gain calculation is incorporated. This approach is motivated from the algorithm proposed by C.T. Lu and H.C. Wang (2004), Ching-Ta Lu ,Chih-Tsung Chen and Kun-Fu Tseng (2010), who proposed data compression scheme that uses psychoacoustic modeling to determine which portions of the audio signal to be removed without loss of sound quality to the human ear. Instead of completely removing the noise (and thereby making it more susceptible to increased distortion and decreased intelligibility) the existing residual noise is masked by exploiting the masking properties (Nathalie Virag 1999) of the human auditory system proposed in the above said references.

The phenomenon of auditory masking has been successfully exploited in the field of wideband audio coding proposed by James D. Jhonston (1988) and by Joachim Thiemann (2001), in which a model of auditory system is used to calculate a spectral masking threshold. The same is used in this subband approach. The psychoacoustic model is based on many studies of human perception. These studies have shown that the average human does not hear all frequencies. Effects due to different sounds in the environment and limitations of the human sensory system lead to facts that are used to cut out unnecessary data in a noisy speech signal.

Finally a perceptual gain factor $g_P(i,j)$ in the frequency domain is obtained as

$$g_{p}(i,j) = \frac{1}{1+\max\left(\sqrt{\frac{\left|D(i,j)\right|^{2}}{T(i,j)}}-1,0\right)}$$

The perceptual gain for enhancement given in the above equation is calculated from the global masking threshold and the estimated noise signal.

Results

The test samples are taken from Speech Enhancement Assessment Resource (SPEAR) database of Center for Spoken Language Understanding (CSLU) and the NOIZEUS database. The sampling rate of the noisy speech samples is 8 kHz and algorithm simulation has been carried out with MATLAB. In this work approximately 50 samples were taken and results of a few are shown below.

Time domain Results:

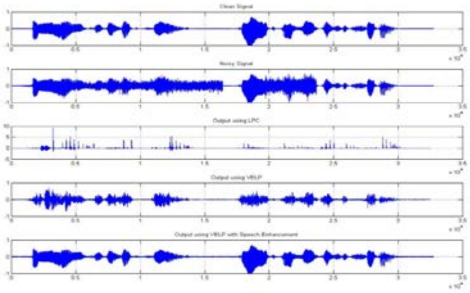

FIGURE 3 TIME DOMAIN PLOTS FOR PROPOSED METHOD (FROM TOP TO BOTTOM CLEAN SIGNAL, NOISY SIGNAL, OUTPUT FROM LPC MODEL WITH NOISY INPUT, OUTPUT USING VELP MODEL WITH NOISY INPUT AND OUTPUT USING VELP WITH SPEECH ENHANCEMENT WITH NOISY INPUT)

Time domain output for the proposed method is shown in figure 3. From the result it is identified that LPC model reproduced the speech signal in time domain with maximum deviation and from VELP it is recovered with noise and distortion. VELP with speech enhancement gives more accurate reproduction of the signal.

Power spectral density for the proposed method is shown in figure 4. The result shown is for one complete sample (Car phone noisy signal). It is illustrated that VELP with speech enhancement is very close to the clean signal and the remaining two methods are much deviated from the clean signal. Comparison of Signal to Noise ration of the proposed method with conventional method is shown in table 1 and Mean Square Error is compared with conventional method (without speech enhancement) in Figure 5. Form the results it is identified that the signal gives better results when it is enhanced before coding.

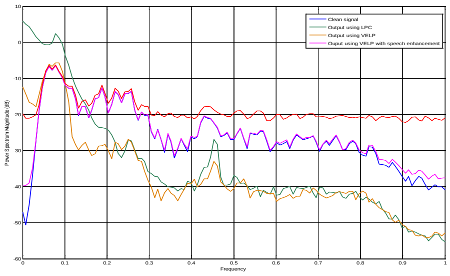

FIGURE 4 POWER SPECTRAL DENSITY FOR THE PROPOSED METHOD (CLEAN SIGNAL, OUTPUT FROM LPC MODEL, OUTPUT USING VELP MODEL AND OUTPUT USING VELP WITH SPEECH ENHANCEMENT)

TABLE 1 COMPARISON OF SNR VALUE OF PROPOSED METHOD WITH EXISTING METHODS

Type of Noisy Signal / Method used	LPC	VELP	VELP with Speech Enhancement
White Stationary 0dB	7.6	19.54	38.34
White bursting 0dB	6.94	12.95	31.81
White bursting 3dB	4.82	15.07	33.49
Car Phone	5.62	11.42	41.26
Cellular	5.39	14.41	25.64
Factory phone	3.12	16.25	36.54
Cockpit noise	5.5	15.42	27.39
F16 factory noise	9.99	11.01	23.12

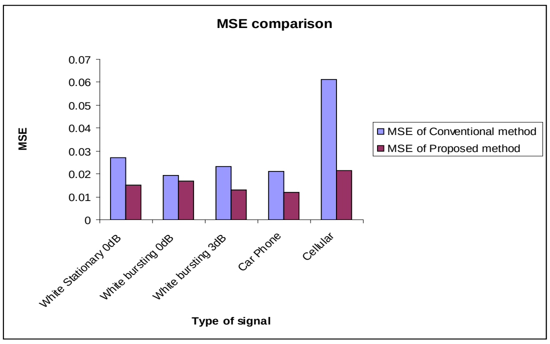

FIGURE 5 MEAN SQUARE ERROR COMPARISON

Conclusion

This proposed method for speech coding with speech enhancement using perceptual speech masking thresholds improves the speech quality by improving Signal to Noise ratio by more than 10 db, which can be accommodated in communication applications to reduce the bitrate with minimum noise level. The Mean square error value is also minimum in the proposed method compared to the conventional LPC and VELP methods. This approach is more suitable for both stationary and non stationary noise environments. Further this method can be implemented in Digital signal processor (TMS 320C6713) for processing real time signals.

REFERENCES

Boll S F, Suppression of acoustic noise in speech signal spectral subtraction, IEEE Trans Acoust Speech Signal Process, 27 (1979) 113-120.

C.T. Lu and H.-C. Wang "Speech enhancement using perceptually constrained gain factors in critical band wavelet packet transform" IEEE Electronics Letters, Vol. 40 No. 6, 2004.

Ching-Ta Lu ,Chih-Tsung Chen and Kun-Fu Tseng, "Speech Enhancement using Perceptual Decision Directed Approach ",Proceedings of IEEE Computer Soceity, Second International Conference on Computer Engineering and Applications, pp 23-27, 2010.

Cohen I, Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging, IEEE Trans Speech Audio Process, 11 (2003) 466-475.

D. Deepa and Shanmugam A, "Enhancement of Noisy speech signal based on Variance and Modified Gain function with PDE preprocessing Technique for Digital Hearing Aids" Journal of Scientific and Industrial Research (JSIR), Vol. 70 PP 332 to 337, 2011.

Deepa D and Shanmugam A, "Speech Enhancement Algorithm Using Sub band Two Step Decision Directed Approach with Adaptive Weighting factor and Noise Masking Threshold " Journal of Computer Science, Vol. 7 No.6: pp. 941-948 (2011).

Ephraim Y & Malah D, Speech enhancement using a minimum mean-square error log-spectral amplitude estimator, IEEE Trans Acoust Speech Signal Process, ASSP-32 (1984) 1109-1121.

James D. Jhonston, "Transform Coding of audio signals using perceptual Noise criteria", IEEE Journal on selected areas in communications, vol. 6, no. 2, 1988.

Joachim Thiemann, Acoustic Noise Suppression for Speech Signals using Auditory Masking effects, Thesis, McGill University Montreal, Canada, July 2001.

Ma J, Hu Y & Loizoub P C, Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions, J Acoust Soc Am, 125 (2009) 3387-3405.

Speech coding overview Jason Woodard, http://www-mobile.ecs.soton.ac.uk/speech_codecs/www.data-compression.com

Sundarrajan R & Philipos C L, A noise estimation algorithm for highly non- stationary environments, Speech Commun, 48 (2006) 220-231.

Virag N, Single channel speech enhancement based on masking properties of the human auditory system, IEEE Trans Speech Audio Process, 7 (1999) 126-137